TP 2: Análisis de Redes Sociales

October 10, 2022 Métodos Numéricos

Grupo 18

Integrante	LU	Correo electrónico
Vekselman, Natán	338/21	`natanvek11@gmail.com`
Arienti, Federico	316/21	`fa.arienti@gmail.com`

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

http://www.exactas.uba.ar

RESUMEN

La descomposición de matrices en autovectores y autovalores aparece en una variedad de aplicaciones donde importa caracterizar el comportamiento de un sistema: en el reconocimiento de imágenes, en el análisis de estabilidad de cuerpos rotantes, en el análisis de riesgo de mercado, y en el análisis de redes — por nombrar algunas —. Desde un punto de vista geométrico, se puede considerar a los autovectores como los 'ejes' de una transformación lineal, en tanto representan una dirección invariante a la transformación, y a los autovalores como los factores por los que esas direcciones se comprimen, estiran o invierten.

En este trabajo implentaremos un método para el cálculo de autovectores y autovalores en matrices con autovalores en módulo diferentes —descontando multiplicidad—, como pueden ser las matrices simétricas definidas positivas, las matrices de conectividad o las matrices laplacianas. El mismo se conoce como *el método de la potencia con deflación*.

A su vez, presentaremos dos aplicaciones concretas de los autovalores y autovectores en el análisis de redes: la medición de centralidad de autovector y conectividad algebráica en la red del *'Club de Karate'* TODO: citar, y el análisis de una red de amigos en *Facebook*, por medio de la construcción de una matríz de similaridad.

Palabras clave: *método de la potencia, deflación de Hotelling, centralidad de autovector, conectividad algebraíca, análisis de componentes principales.*

CONTENIDOS

1. Método de la potencia con deflación

1.1. **Introducción teórica.**
El método de la potencia con deflación permite aproximar un subconjunto de los autovalores y autovectores asociados a una matríz. Si la misma satisface que todos sus autovalores (descontando multiplicidad) son diferentes en módulo, entonces permite aproximar el conjunto entero. El método se compone de dos partes:

Método de la potencia: el método de la potencia, *Power method* ó *Power iteration*, es una técnica iterativa para aproximar el autovector asociado al autovalor de módulo máximo de una matríz cuadrada que satisfaga esta característica —es decir, tenga un autovalor dominante—, a partir de la aplicación de sucesivos productos matriciales, descriptos por la siguiente relación de recurrencia:

$$b_0 \text{ es un vector aleatorio unitario}$$

$$b_{k+1} = \frac{\mathbf{A} b_k}{||\mathbf{A} b_k||} \tag{1}$$

donde $||\cdot||$ es una norma vectorial.

Se puede demostrar [citar] que, bajo las condiciones mencionadas, si b_0 no es ortogonal al autovector asociado al autovalor dominante en módulo de $\mathbf{A}$, b_k convergerá a éste. Lo que es más, se podrá aproximar el autovalor dominante por medio del coeficiente de Rayleigh:

$$\lambda_{max} = \frac{b_k^t \ \mathbf{A} \ b_k}{b_k^t \ b_k} \tag{2}$$

Método de la deflación: el método de la deflación, por su parte, corresponde a la transformación de la matríz inicial $\mathbf{A}$ por una matríz $\mathbf{B}$ que comparta autovalores con $\mathbf{A}$, salvo por el autovalor dominante, que se reemplazará por el autovalor 0. Existen numerosos métodos de deflación, entre ellos la deflación de Hotelling [citar] y la deflación de Wielandt [citar] . [completar con una explicación del método que usemos.]

1.2. **Implementación.**
Procedemos a detallar una posible implementación para ambos métodos. Para no complicar la implementación, nos restringiremos a las matrices cuyos autovalores asociados son reales. Definimos entonces:

$$
\begin{aligned}
deflacion: \ & \textit{matriz}\text{<n} \times \text{n>} \ \mathbf{A} \ \times \ \textit{nat} \ \text{q} \ \times \ \textit{nat} \ \text{k} \ \times \ \textit{real} \ \text{t} \\
& \longrightarrow \ \textit{vector}\text{<q>} \ \times \ \textit{matriz}\text{<n} \times \text{q>}
\end{aligned}
$$

$$
\begin{aligned}
\textit{potencia}: \ & \textit{matriz}\text{<n} \times \text{n>} \ \mathbf{A} \ \times \ \textit{nat} \ \text{k} \ \times \ \textit{real} \ \text{t} \\
& \longrightarrow \ \textit{real} \ \times \ \textit{vector}\text{<n>}
\end{aligned}
$$

donde n es un natural, $\mathbf{A}$ tiene q autovalores reales dominantes en módulo, $0 < q \leq n$, k representa el número máximo de iteraciones a realizar y $0 \leq t$ representa la tolerancia de error a partir de la que una solución se considera aceptable.

Proponemos el siguiente algoritmo para $Deflacion$:

```
proc deflacion(in A: matriz<n, n>, in q: Nat, in k: Nat, in t: Real) {
    // TODO
}
```

Algoritmo 1. Pseudocódigo para el método de la deflación.

El mismo retornará un vector con los q autovalores de $\mathbf{A}$, ordenados descendientemente por tamaño, y una matriz cuyas columnas corresponden, respectivamente, a los autovectores unitarios asociados a los q autovalores. [explicar un toque aspectos relevantes a su implementacion.]

Potencia, por su parte, retornará el autovalor de $\mathbf{A}$ máximo en módulo y su autovector asociado. Proponemos el siguiente algoritmo:

```
proc potencia(in A: matriz<n, n>, in k: Nat, in t: Real) {
    // TODO
}
```

Algoritmo 2. Pseudocódigo para el método de la potencia.

[explicar aspectos relevantes.]

1.3. **Evaluación cuantitativa.** Procederemos a evaluar nuestra implementación del método de la potencia con deflación en C++ acorde a los algoritmos propuestos.

Error relativo Medimos el error $|\mathbf{AV} - \mathbf{V\Lambda}|_1$ en función de la cantidad de iteraciones k para 300 instancias de matrices $\in \mathbb{R}^{25\times25}$ generadas aleatoriamente, donde $\mathbf{V}$ y $\mathbf{\Lambda}$ representan

—respectivamente— las matrices aproximadas de autovectores y autovalores de $\mathbf{A}$, tal que $\mathbf{A}\mathbf{V}_i \approx \mathbf{\Lambda}_{ii}\mathbf{V}_i\ \forall i:\ 1\ ...\ n$. En total, obtuvimos [XXX] mediciones[1].

METODOLOGÍA. Se calculó $\mathbf{\Lambda}, \mathbf{V} = Deflacion(\mathbf{A},\ 25,\ k,\ 0)$ y se midió el error relativo para cada una de las matrices sobre cada valor de k en en el intervalo $(0, 1e6)$, de a saltos de $1e3$.

Cada caso se generó a través de uno de los siguientes tres procedimientos[2]:

1) *Matrices Diagonales*: Se generaron cien matrices diagonales con veinticinco autovalores en el rango $[-1e6,\ 1e6]$ tal que ningun autovalor compartiera valor absoluto en módulo con ningún otro, dados signos diferentes. Los mismos se eligieron a partir del rng PCG64 de numpy para evitar distribuciones particulares que pudieran influir en la variabilidad de los autovalores. Notar que los autovalores de una matríz diagonal son los elementos de su diagonal.

2) *Matrices Diagonalizables*: Se generaron cien matrices diagonalizables $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^t$ donde cada matríz $\mathbf{D}$ se generó a partir de la metodología (1.) y $\mathbf{Q}$ es un matríz de householder $\mathbf{Q} = \mathbf{I} - 2uu^t$, con u un vector aleatorio —generado a partir del algoritmo random.rand() de numpy— tal que $||u||_2 = 1$.

3) *Matrices Simétricas Definidas Positivas*: Se generaron cien matrices simétricas definidas positivas de enteros de la siguiente manera: se generó la matríz $\mathbf{B}$ con el algoritmo random.randint() de numpy y se definió la matríz $\mathbf{A} = \mathbf{B}\mathbf{B}^t$.

RESULTADOS. [experimento]

[1]El script asociado se puede encontrar en $./experimentos/error\_potencia.py$

[2]Se utilizó un valor semilla para facilitar la reproductibilidad.

2. Análisis: Club de Karate

2.1. **Contexto.** La red del *Club de Karate* es parte de una investigación antropológica [citar] que estudió las relaciones 'políticas' entre los miembros de un club universitario. La misma se realizó durante el desarrollo de un conflicto que terminó por dividir al grupo. La red busca modelar el flujo de información entre sus integrantes por medio de la tripla $(V,\ E,\ C)$, donde V denota el conjunto de individuos, E refiere a un grafo no direccionado cuyos ejes representan los vínculos entre los miembros, y C define la fuerza de estas relaciones —lo que se podría pensar como ponderaciones sobre los ejes de E—.

La investigación culmina con una demostración de la capacidad del modelo para predecir la división del grupo por medio del algoritmo de *labeling* de 'flujo máximo - corte mínimo' de Ford y Fulkelson [citar] .

En este análisis utilizaremos sólo la representación matricial de E para evaluar la importancia de los distintos miembros en la red, y la matríz laplaciana asociada para evaluar el uso de autovectores como predictores de la división del grupo.

2.2. **Centralidad de Autovector.** La centralidad de autovector es una medida utilizada en el análisis de redes para medir la 'importancia' de los nodos que la componen, relativa a la importancia de sus conexiones. Dada una matríz de conectividad $\mathbf{W}$, y un vector inicial x, se define:

$$x'_i = \sum_j \mathbf{W}_{ij} x_j \tag{3}$$

o lo que es equivalente:

$$x' = \mathbf{W}x \tag{4}$$

Intuitivamente, se puede pensar que la importancia de cada nodo es proporcional a la suma de las importancias de sus vecinos [citar] .

Vemos que la equación (5.) es similiar a (1.). Se puede demostrar [citar] que la aplicación sucesiva de este método —tal que $x_{k+1} = \mathbf{W}x_k$— permite estimar el autovector de módulo máximo, y este será positivo:

$$\lambda x = \mathbf{W}x \tag{5}$$

Podemos ver que la aplicación del metodo de la potencia con deflación sobre la matríz de conectividad asociada al grafo E resulta en el siguiente autovalor y autovector asociado:

[presentar vector.]

Vemos que el nodo '1' y el nodo '34' son los más centrales. Esto no es casualidad, la red del 'Club de Karate' está armada para tener a las dos figuras centrales del conflicto en cada extremo —el instructor de karate y el presidente del club—, para satisfacer la especificación del algoritmo de 'labeling' que utiliza.

2.3. **Autovectores de la matríz laplaciana.** La matríz laplaciana $\mathbf{L} = \mathbf{D} - \mathbf{W}$, donde $\mathbf{D}$ es una matríz diagonal con elementos $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ y $\mathbf{W}$ es la matríz de conectividad asociada al grafo E, sirve para medir distintas propiedades de una red. En particular, el mínimo autovalor no nulo —llamado conectividad algebraica o valor de Fiedler— permite establecer un criterio sobre el que particionar la red en dos, donde el autovector asociado designará la pertenencia de un nodo a una u otra partición acorde a su signo.

Procedemos a analizar qué autovector de la matríz laplaciana asociada a E permite predecir mejor la división que ocurrió en el club. Para ello, medimos el valor absoluto de la correlación entre cada autovector y el vector que indica el grupo:

[presentar resolución.]

3. Análisis: Facebook

3.1. Contexto.

3.2. Matríz de similaridad, matríz de atributos.

3.3. Comparación con la red original.

3.4. Optimización.

3.5. PCA.

4. Conclusiones

5. Apéndice